International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

# UNET-Based Terrain Classification for Enhanced Environmental and Geospatial Applications

***[1]Dr. Rasmi A, [2]Aishwarya L, [3]Chaitra R, [4]Sharath S, [5]Syeda Samha***

[1]Guide, Associate Professor, RR Institute Of Technology, Bengaluru
[2,3,4,5] RR Institute of Technology, Bengaluru, India
[2]aishwaryaloganathan22@gmail.com, [3]chaitra051103@gmail.com, [4]sharathsrao4529@gmail.com, [5]worksamha@gmail.com

**ABSTRACT—**

Terrain classification involves identifying and categorizing land cover types based on characteristics like elevation, slope, vegetation, and land usage. Terrain classification is important to various disciplines such as remote sensing, geology, environmental studies, urban development, agriculture, and military surveys. Maintaining a highly accurate, detailed, and not over-smoothed DTM is essential for advancing technologies. Several classical DSM filtering algorithms and recent deep learning-based methods have been developed. However, most are multi-step processes and unable to maintain sharp terrain slopes, particularly in terraced terrain. In this work, a deep learning-based semantic segmentation model using UNET is proposed for classifying primary landform categories using 30m-resolution DEM data. The aim is to evaluate the capability of deep learning to classify recurring landform categories.

**Keywords**: UNET, Terrain Classification, DEM, Deep Learning, Semantic Segmentation, Landform Mapping, Remote Sensing.

## I. INTRODUCTION

Terrain classification is the classification of various kinds of terrain or land cover of an area against some characteristics such as elevation, slope, vegetative cover, and land use. It is important to various disciplines such as remote sensing, geology, environmental studies, urban development, agriculture, and military surveys. The Earth’s surface, devoid of vegetations and man-made structures like buildings and roads, provides a basis for Digital Terrain Models (DTMs), which are crucial in surveying, engineering, disaster management, and land cover mapping. Therefore, maintaining a highly accurate, detailed, and not over-smoothed DTM is essential for advancing technologies.

DTMs may be produced directly from terrain observations or derived from Digital Surface Models (DSMs), which in turn can be created through active sensing methods like laser scanning and radar interferometry, or from optical stereo images. Extracting a DTM from a DSM involves removing above-ground objects and interpolating voids with suitable height data. While classical DSM filtering algorithms exist, most are multi-step and depend on predefined thresholds. These methods often fail to preserve sharp terrain slopes, particularly in terraced landscapes.

Landforms are essential components of the natural landscape and influence the ecological environment and distribution of natural resources. Mapping these landforms at large scales has been historically difficult due to limited systematic data. However, advances in remote sensing and GIS technology have enabled the use of satellite imagery and Digital Elevation Models (DEMs) for semi-automatic and automatic landform classification, making it feasible for broader applications.

DEMs simulate the Earth’s surface through elevation data and allow the extraction of parameters such as aspect, slope, and curvature, from which landform classes can be inferred. Terrain classification methods using DEMs typically rely on either pixel-based or object-based approaches. Pixel-based methods assign landform labels based on elevation thresholds but often lack spatial consistency. Object-based techniques, in contrast, segment images into meaningful morphological units and reflect more accurate geomorphological features, though they struggle with dynamically changing landforms.

With growing computational power and data availability, deep learning has become a viable alternative. Deep learning models automatically learn hierarchical features without domain-specific programming. Researchers have applied deep learning in geomorphology for landform detection using imagery and DEM data. The U-Net model, a popular semantic segmentation architecture, has been adopted to classify loess landforms, showing the potential of deep learning in terrain classification.

Despite promising results, semantic segmentation models were originally designed for natural images with clear object boundaries, unlike the continuous variation in DEM-based terrain data. Therefore, their application to geomorphological data requires further evaluation. In this work, we have developed a deep learning-based semantic segmentation model using the U-Net architecture for classifying primary landform categories from 30-meter-resolution DEM data at the pixel level. This study aims to test the suitability of deep learning in classifying repeating landform types across diverse terrains.

## II. LITERATURE SURVEY

[1]. In the paper *"Proprioception Is All You Need: Terrain Classification for Boreal Forests"* the authors Damien LaRocque, William Guimont-Martin, David-Alexandre Duclos, Philippe Giguère, and François Pomerleau introduced BorealTC, a dataset for proprioceptive-based terrain classification (TC) recorded using a Husky A200 robot. The dataset contains 116 minutes of IMU, motor current, and wheel odometry data across boreal forest terrains like snow, ice, and silty loam. The authors tested both a CNN and a new Mamba-based state-space model on the dataset. They found CNNs perform better on individual datasets, while Mamba performs better when trained on a merged dataset. Additionally, they demonstrate that Mamba's learning ability scales better than CNNs with increased data volume and that merged datasets yield latent spaces interpretable by terrain characteristics.

[2]. In the paper *"MARC-Net: Terrain Classification in Parallel Network Architectures Containing Multiple Attention Mechanisms and Multi-Scale Residual Cascades"* the authors Xiangsuo Fan, Xuyang Li, Chuan Yan, Jinlong Fan, Ling Yu, Nayi Wang, and Lin Chen introduced a parallel network architecture land-use classification model based on a composite multi-head attention mechanism and multiscale residual cascade, known as MARC-Net. The parallel framework explores features extracted by grouped spectral embedding and introduces a multi-head attention mechanism to enable semantic features to express more subspaces while capturing all spatial interrelationships. A multiscale residual cascade CNN fully utilizes the feature information of fused features at multiple scales, enhancing network representational capacity. These multi-head attention features are fused with CNN features and downgraded via fully connected layers for pixel-level multispectral image classification. The model achieves a total precision of 97.22%, outperforming the Vision Transformer (ViT) which scored 95.08%. The study also emphasizes changes in forest land cover over time in the research area, noting shifts in coverage from 2017 to 2021.

[3]. In the paper *"Human-Aided Online Terrain Classification for Bipedal Robots Using Augmented Reality"* the authors Zahraa Awad, Celine Chibani, Noel Maalouf, and Imad H. Elhajj presented an augmented reality-based online training system for enhancing real-time terrain classification by humanoid robots. The real-time prediction model is based on data from four sensors (current, inertial, position, and force) of the NAO humanoid robot. They compare prediction performance using Stochastic Gradient Descent, Passive Aggressive classifier, and Support Vector Machine. The models are trained online by manually entering the correct terrain type to enhance prediction accuracy. An augmented reality interface is used to display robot diagnostics and allow manual corrections to predicted terrain types. This feedback improves both classification results and the data collection process. Experimental results indicate that the Passive Aggressive classifier is the most accurate among the evaluated models, achieving 81.4% accuracy.

[4]. In the paper *"Extracting Terrain Texture Features for Landform Classification Using Wavelet Decomposition"* the authors Yuexue Xu, Shengjia Zhang, Jinyu Li, Haiying Liu, and Hongchun Zhu proposed terrain texture classification using the texture feature, which is presented to express landform spatial differentiation and homogeneity. Using the ALOS World 3D-30m (AW3D30) DEM and taking common landforms of the southwest Tibet Plateau as examples, the discrete wavelet transform (DWT) is implemented to analyze the multiscale structural characteristics of the terrain texture. Through structural indices of reconstructed texture images, the optimal decomposition scale of DWT is determined. The wavelet coefficients and the wavelet energy entropy are then used as texture features. Finally, a random forest (RF) algorithm is employed to identify landform. The experimental results show that the DWT texture feature achieves higher classification precision, with improvement of about 11.8% over the gray co-occurrence matrix (GLCM).

[5]. In the paper *"Self-Supervised Visual Terrain Classification From Unsupervised Acoustic Feature Learning"* the authors Jannik Zürn, Wolfram Burgard, and Abhinav Valada introduced a new terrain classification paradigm using an unsupervised proprioceptive classifier trained on vehicle-terrain interaction sounds to self-supervise an exteroceptive classifier for pixel-wise semantic image segmentation. For this purpose, they first learn a discriminative embedding space for vehicle-terrain interaction sounds from audio clip triplets constructed based on visual features of the corresponding terrain patches and cluster the resultant embeddings. Then, these clusters name the patches of the visual terrain by projecting the robot's paths traveled into the images of the cameras. Lastly, they train their semantic segmentation network in a weakly supervised fashion using the sparsely labeled images. They provide extensive quantitative and qualitative results showing that their proprioceptive terrain classifier outperforms the state-of-the-art among unsupervised approaches and that the self-supervised exteroceptive semantic segmentation model performs similarly to supervised learning with human-labeled data.

## III. PROPOSED WORK

The proposed terrain classification system is a deep learning-based architecture that enables accurate, real-time identification of diverse terrain types across varying environmental conditions.

This framework is designed to overcome the limitations of traditional manual classification and computational inefficiencies of existing models by incorporating advanced techniques like UNET convolutional networks, sensor fusion, and real-time online feedback mechanisms. It leverages high-resolution satellite imagery and digital elevation models to extract meaningful terrain features. The system is optimized to handle noise, illumination changes, and seasonal variations, ensuring robust performance in complex, real-world environments. This makes it particularly suitable for applications in geospatial and environmental monitoring scenarios, including disaster risk assessment, land use planning, and autonomous navigation.

1. Core Classification Framework – UNET Architecture

At the center of the system is the UNET model, a convolutional neural network (CNN) designed specifically for semantic segmentation tasks. It features a symmetric encoder-decoder architecture with skip connections that preserve spatial resolution while learning contextual features. The encoder captures terrain features using convolution and max-pooling layers, while the decoder performs up sampling and reconstruction to generate segmentation masks.

This structure enables precise pixel-level classification of terrain types such as forest, hill, desert, and water bodies. The skip connections play a critical role by reintroducing fine-grained spatial information lost during down sampling, improving boundary delineation. The model's flexibility allows it to be extended with attention mechanisms and residual blocks to further boost classification accuracy and generalization.

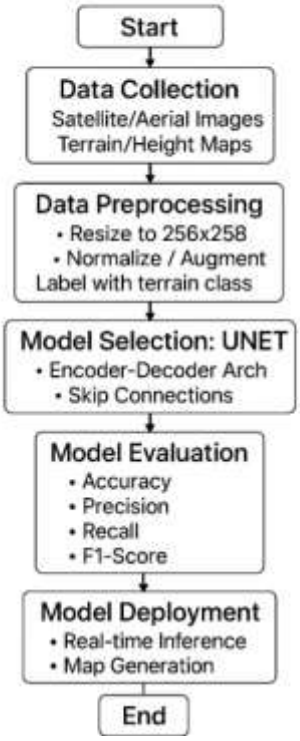



Fig 1. UNET Architecture Overview

2. Image Preprocessing Pipeline

The input images, sourced from satellite or aerial imagery, are standardized to a uniform resolution of 256x256 pixels to ensure consistency. The preprocessing phase includes normalization and augmentation (e.g., flipping, rotation) to increase model generalization. Augmentation helps in creating a more diverse training dataset, which reduces overfitting and improves model robustness against transformations. Images are labelled into categories like mountain, grassland, tundra, water, and desert, and then partitioned into training, validation, and testing subsets to ensure unbiased performance evaluation. Additionally, noise reduction filters and histogram equalization are applied to enhance image quality before feeding into the model.

3. Sensor Fusion for Robustness

To enhance classification accuracy under variable conditions (e.g., lighting changes, sensor noise), the system integrates sensor fusion techniques. Multispectral images, texture features, and elevation data (DEM) are combined to provide a rich, context-aware input to the classification engine. This fusion not only leverages complementary data modalities but also mitigates the limitations inherent to each sensor type alone. For instance, while multispectral images capture vegetation indices, DEM provides elevation and slope information crucial for distinguishing similar spectral signatures in complex terrains. This multi-source data fusion significantly increases the model's resilience in scenarios like shadowed regions or seasonal snow cover.

4. Training Process and Optimization

The model is trained using supervised learning with categorical cross-entropy as the loss function. A batch size of 32, learning rate of 1e-4, and 40 epochs were found optimal for the dataset of 5000 terrain images. The training process includes early stopping based on validation loss to prevent overfitting. Regularization techniques like dropout and L2 weight decay are applied throughout the network to improve generalization. The system also uses learning rate scheduling to dynamically adjust learning rates and accelerate convergence. Performance metrics such as accuracy, precision, recall, and F1-score are computed per terrain class to provide detailed insight into the classification quality.

5. Output and Visualization

The final output consists of high-resolution land cover maps that visually highlight different terrain zones using distinct labels and color codes. These maps are compatible with standard GIS platforms and support overlays with other environmental data layers. Visualization tools enable zoom, pan, and attribute queries, aiding in detailed spatial analysis. The classified maps assist stakeholders in applications such as disaster risk assessment, precision agriculture, habitat conservation, and urban planning. The system can also generate temporal change detection maps to monitor landscape evolution over time.

6. Scalability and Deployment

The proposed system is lightweight enough for deployment on resource-constrained platforms such as mobile robots, UAVs, or remote sensing drones. Its modular architecture allows seamless integration with different sensor types and computing hardware, from embedded GPUs to cloud-based servers. The inference engine supports batch and real-time processing modes, enabling flexible usage scenarios. Scalability is achieved via distributed training and data parallelism to handle very large datasets. This makes the system suitable for both localized field operations and large-scale geospatial surveys.

## IV. METHODOLOGY

The proposed terrain classification system is a deep learning-based framework designed for accurate and efficient identification of diverse terrain types across varied environmental conditions. This system leverages the UNET architecture for semantic segmentation and integrates preprocessing, dataset engineering, training, and optimization to support real-time geospatial and environmental applications. The goal is to enable automation of land cover mapping for decision-making in environmental monitoring, disaster response, and agricultural management.

1. Data Collection and Preprocessing

The input data, including satellite or aerial images, undergoes a preprocessing pipeline to ensure uniformity and efficient processing. Initially, the images are resized to a fixed resolution of 256×256 pixels. A labelled dataset is curated where each image is annotated with terrain class labels such as water, hill, forest, grassland, desert, mountain, and tundra. These are then split into training, validation, and testing subsets to facilitate systematic training and evaluation of the model. Additionally, images undergo noise filtering and augmentation techniques such as rotation and horizontal flipping to improve data diversity and model generalization.

2. Image Resizing and Standardization

Image resizing ensures uniformity in data dimensions and enables compatibility with neural network input requirements. Resizing also improves computational efficiency during training and inference while preserving critical spatial and contextual information. Techniques such as padding and cropping are used to retain the aspect ratio. The chosen resizing resolution is 256x256 pixels, which balances classification accuracy and resource efficiency. The interpolation method used for resizing plays a significant role in preserving feature detail; bilinear and bicubic interpolation are selected based on the image quality needed. Resizing is implemented automatically within the preprocessing pipeline using integrated deep learning libraries to ensure consistency.

3. Dataset Composition

The dataset comprises 5000 image sets, each a 512×512 pixel crop representing a combination of a terrain map, a height map, and a segmentation map. Terrain maps are coloured based on land type and include relief shading, while height maps encode elevation values using 16-bit grayscale data. Segmentation maps are generated using unsupervised clustering techniques and categorized into seven terrain classes—water, grassland, forest, hills, desert, mountain, and tundra—with each category assigned distinct RGB colour values. Median filtering is used to eliminate noise and smooth segmentation boundaries. Variability is introduced by adjusting crops based on latitude to maintain scale and introducing randomization in segmentation to improve robustness.

4. Core Classification Engine – UNET Architecture

The UNET model employed features an encoder-decoder structure with skip connections, allowing for the preservation of spatial resolution and hierarchical feature extraction. The encoder captures contextual features via convolutional and pooling layers, while the decoder reconstructs segmentation masks through up-sampling and concatenation with encoder features. This design enables fine-grained, pixel-level terrain classification with high accuracy. Each level in the encoder reduces the spatial size while increasing the number of feature maps. The decoder applies transposed convolutions to restore original resolution, supported by skip connections that bridge the encoder and decoder paths to retain fine-grained edge and boundary features.

5. Advanced Variants – ResUNET and Attention UNET

To improve the learning ability and training convergence of the standard UNET, enhancements like Res-UNET and Attention-UNET are integrated. Res-UNET uses residual connections to simplify training and improve feature learning, reducing the chances of vanishing gradients in deep architectures. Attention-UNET introduces attention gates that automatically highlight relevant features and suppress irrelevant ones. These gates improve accuracy while reducing model complexity and computational overhead. The attention mechanism dynamically focuses the model's response on target structures of varying sizes and shapes, while ignoring less informative regions, thereby improving the interpretability and precision of the segmentation.

6. Training Strategy and Hyperparameters

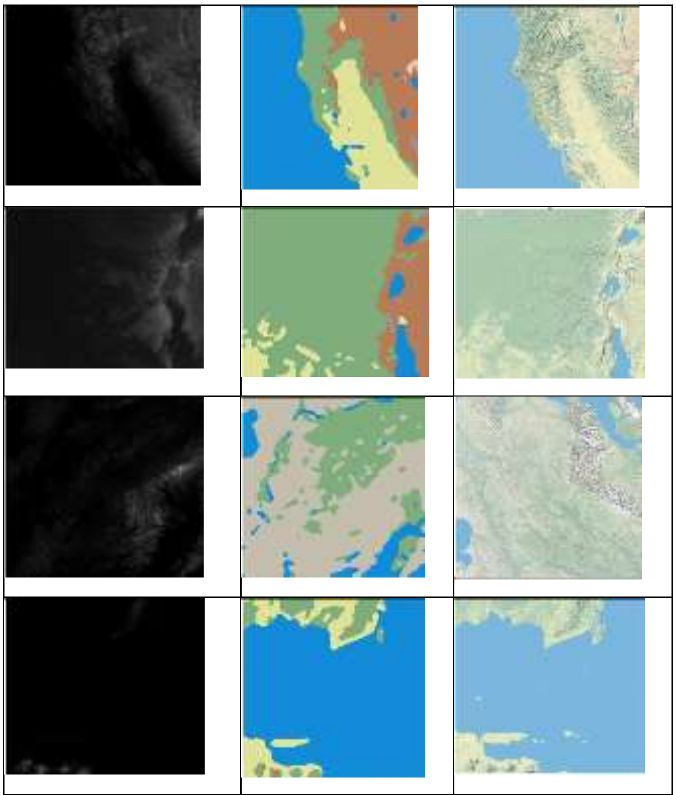
The model is trained using supervised learning with a categorical cross-entropy loss function. Training parameters include a batch size of 32, a learning rate of 1e-4, and 40 training epochs. Regularization techniques such as L1, L2, and dropout are used to prevent overfitting and enhance generalization. The ReLU activation function is used in all convolution layers, and max pooling is applied in the encoder for down-sampling. Up-sampling in the decoder is achieved via transposed convolution layers. Model training is monitored using validation loss, and early stopping is employed to avoid overfitting. The performance of the final trained model is evaluated using metrics like accuracy, precision, recall, and F1-score across all terrain classes.

7. Output and Visualization

The output of the system is a high-resolution terrain classification map, where each pixel is tagged with its corresponding terrain class. These maps are presented using color-coded overlays that correspond to the defined class labels. The segmentation outputs are used to support tasks such as land use monitoring, disaster zone mapping, and ecological trend analysis. The visualized outputs are compatible with standard GIS tools for further integration and analysis. The classification results are compared with ground truth labels to validate model performance, and confusion matrices are generated to identify per-class accuracy and misclassification patterns.

8. Scalability and Real-Time Application

The architecture is optimized for deployment in real-time or resource-limited environments such as mobile robotic platforms, UAVs, and edge devices. Despite its complexity, the lightweight model variations such as UNET-lite or pruned Res-UNET reduce memory footprint and inference time. Furthermore, AR-based interfaces are considered for real-time user feedback, where incorrect predictions can be manually corrected, feeding into a continuous learning loop for future model refinement.



| (a) | (b) | (c) |
|---|---|---|

Fig 2. a) Masked maps b) CNN based Classified maps c) Actual maps

Satellite imagery or aerial photographs are collected and preprocessed to ensure consistency and facilitate efficient processing. The input images are resized to a standardized format of 256×256 pixels, enabling uniformity in data dimensions for subsequent processing stages. A labelled dataset is curated, comprising images annotated with ground truth labels corresponding to different terrain classes such as water, hill, forest, grassland, desert, mountain, and tundra. The dataset is divided into training, validation, and testing subsets to support systematic model development.

Image resizing ensures uniformity in data dimensions and compatibility with deep learning input requirements. The selected size of 256×256 pixels balances classification accuracy with computational efficiency. Techniques such as cropping or padding are applied to maintain the aspect ratio during resizing. The resizing operation is automated and integrated into the data pipeline using tools like OpenCV, minimizing manual intervention and supporting batch processing.

The dataset comprises 5000 image sets, each representing a 512×512 pixel crop of Earth, including a terrain map, a height map, and a segmentation map. The segmentation maps are generated using unsupervised clustering and classification techniques and are assigned one of seven terrain classes. Relief shading and color coding are used for visual differentiation. Maps are further processed using median filtering to remove noise and enhance segmentation clarity.

The UNET model forms the core of the system. It follows an encoder-decoder architecture with skip connections that preserve spatial information during the transformation from input to segmentation mask. The encoder consists of convolutional and pooling layers that extract features, while the decoder reconstructs the image via up-sampling and concatenation with features from the encoder. The UNET architecture is specifically selected for its ability to deliver pixel-wise semantic segmentation with high resolution and precision.

The training process is performed using a supervised learning approach. The model learns to associate input images with their corresponding terrain classes by minimizing a categorical cross-entropy loss function. Training parameters include a batch size of 32, a learning rate of 1e-4, and 40 training epochs. Regularization methods such as dropout, L1, and L2 are used to prevent overfitting. Model performance is monitored using validation metrics including accuracy, precision, recall, and F1-score.

Once trained, the model is tested on unseen data (testing subset). It generates terrain class labels for each pixel in the input image, creating a segmentation map. These predictions are compared against the ground truth labels to assess model reliability and classification performance. The output includes terrain maps useful for applications in environmental monitoring, land management, and urban planning.

In total, 7 terrain categories were defined for segmentation and associated with representative color as follows:

- (17, 141, 215): Water
- (225, 227, 155): Grassland
- (127, 173, 123): Forest
- (185, 122, 87): Hills
- (230, 200, 181): Desert
- (150, 150, 150): Mountain
- (193, 190, 175): Tundra

A Flask-based web interface is used to deploy the trained model. The user uploads an input image through the web interface, which is processed through the backend where the UNET model performs classification. The resulting segmented image is displayed back to the user, marking different terrain regions with distinct colors. This supports real-time terrain visualization through a user-friendly frontend.

Fig 3. Command Prompt

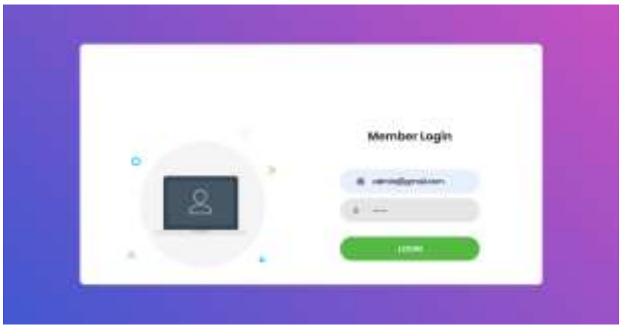



Fig 4. Login page


Fig 5. Output Image generated

The system is implemented using Python and relies on machine learning libraries like TensorFlow, PyTorch, OpenCV, and NumPy. Flask is used to build the web interface for model deployment, while Google Colab facilitates model training and visualization during development.

## V. RESULTS AND DISCUSSION

The terrain classification system based on UNET architecture was implemented and tested successfully. The process involved launching the main Python script, navigating to the system's local web interface (https://127.0.0.1:5000), logging in with valid credentials, uploading satellite or aerial imagery, and observing the classified output image.

Step-by-step operation:

1. Execute the main Python file using the command python main.py.
2. Navigate to the system interface using a browser.
3. Input the user credentials to access the platform.
4. Select and upload a terrain image for classification.
5. The system processes the input and generates an output image that labels the terrain classes accordingly.

The output images display segmented terrain regions where each pixel is annotated with its respective terrain class such as water, hill, forest, grassland, desert, mountain, or tundra.

The presented terrain classification framework demonstrates the effectiveness of leveraging the UNET architecture for accurately categorizing diverse terrain types based on input imagery resized to a standardized format of 256×256 pixels. Through extensive experimentation and evaluation, the framework achieved high classification accuracy across various terrain categories.

The UNET model efficiently extracted discriminative features and spatial relationships necessary for terrain classification. Training the network on labelled datasets enabled the system to recognize complex patterns and variations characteristic of each terrain type. The classification output exhibited strong boundary delineation capabilities and high semantic precision in identifying landscape regions.

The system also proved to be scalable and adaptable for large-scale datasets, making it suitable for real-world applications in environmental monitoring, urban planning, disaster management, and natural resource analysis. Future enhancements may include integrating data modalities such as LiDAR or hyperspectral imagery to further improve accuracy and robustness.

## VI. FUTURE SCOPE

The UNET-based terrain classification system can be further enhanced by integrating additional data modalities such as LiDAR point clouds, hyperspectral imagery, and synthetic aperture radar (SAR). These sources provide richer spatial and spectral information, which can significantly improve classification accuracy and robustness across complex and heterogeneous landscapes. Hybrid deep learning architectures that combine RGB imagery, height maps, and terrain texture data using attention mechanisms and residual connections may also be explored to improve generalization and scalability.

Another promising area involves improving model interpretability and adaptability. Attention mechanisms and uncertainty estimation techniques can help visualize the decision-making process, allowing experts to better validate and trust the output. Additionally, incorporating an augmented reality (AR)-based user feedback interface could enable real-time corrections and adaptive learning. This would allow the model to refine itself continuously through field-based inputs and maintain high accuracy in new or evolving terrains.

Finally, future work should focus on making the system lightweight and suitable for deployment on mobile platforms such as drones or ground robots. This includes optimizing the model for low-latency, real-time inference and ensuring compatibility with embedded systems. Consideration must also be given to ethical deployment, data privacy, and standardization to ensure responsible use and integration into geospatial workflows. By addressing these aspects, the system can be scaled and applied in disaster response, ecological monitoring, and sustainable land-use planning.

## VII. CONCLUSION

This study presented a deep learning-based terrain classification system utilizing the UNET architecture for precise semantic segmentation of satellite and aerial imagery. The system was designed to address the limitations of manual classification methods by automating the identification of diverse terrain types, including water, forest, hills, deserts, and mountains. Through systematic preprocessing, structured dataset preparation, and the use of convolutional neural networks, the model achieved pixel-level accuracy and maintained spatial integrity during classification.

The implementation included key enhancements such as image normalization, augmentation, and regularization techniques, which significantly improved model performance and generalization. Variants like Res-UNET and Attention UNET further strengthened the system's ability to focus on relevant terrain features while reducing noise. The results demonstrated that the model was effective not only in static evaluation but also in real-time prediction through a user-friendly web interface, making it practical for geospatial analysis tasks.

Overall, the proposed system proves to be a scalable and robust solution for environmental monitoring and land-use planning. Its lightweight architecture makes it deployable on resource-constrained platforms such as UAVs and embedded systems. With future enhancements, including multimodal data integration and real-time adaptive learning, the system can serve as a valuable tool for sustainable natural resource management and decision support in dynamic environmental conditions.

## VIII. REFERENCES

[1] LaRocque, Damien, William Guimont-Martin, David-Alexandre Duclos, Philippe Giguère, and François Pomerleau. *"Proprioception Is All You Need: Terrain Classification for Boreal Forests."* arXiv preprint arXiv:2403.16877 (2024).

[2] Kaur, Sukhjeet, and Sanjay Madaan. *"A hybrid UNet based approach for crop classification using Sentinel-1Bs synthetic aperture radar images."* Multimedia Tools and Applications (2024): 1–30.

[3] Fan, Xiangsuo, Xuyang Li, Chuan Yan, Jinlong Fan, Ling Yu, Nayi Wang, and Lin Chen. *"MARC-Net: Terrain classification in parallel network architectures containing multiple attention mechanisms and multi-scale residual cascades."* Forests 14, no. 5 (2023): 1060.

[4] Zhang, Guanjin, Siti Nur Aliaa binti Roslan, Ci Wang, and Ling Quan. *"Research on land cover classification of multi-source remote sensing data based on improved U-net network."* Scientific Reports 13, no. 1 (2023): 16275.

[5] Chen, Yang, Yao Wu, Wei Zeng, and Shaoyi Du. *"Kinematics Model Estimation of 4W Skid-Steering Mobile Robots Using Visual Terrain Classification."* Journal of Robotics (2023).

[6] Awad, Zahraa, Celine Chibani, Noel Maalouf, and Imad H. Elhajj. *"Human-Aided Online Terrain Classification for Bipedal Robots Using Augmented Reality."* In 2022 IEEE International Conference on Robotics and Biomimetics (ROBIO), pp. 1352–1358. IEEE, 2022.